JOURNAL OF MULTIVARIATE ANALYSIS 57, 175–190 (1996)
ARTICLE NO. 0028

Order Determination for Multivariate Autoregressive Processes Using Resampling Methods

CHANGHUA CHEN AND RICHARD A. DAVIS*

Colorado State University

AND

PETER J. BROCKWELL*

Royal Melbourne Institute of Technology, Melbourne, Australia

Let $\mathbf{X}_1, ..., \mathbf{X}_n$ be observations from a multivariate AR(p) model with unknown order p. A resampling procedure is proposed for estimating the order p. The classical criteria, such as AIC and BIC, estimate the order p as the minimizer of the function

$$\delta(k) = \ln(|\hat{\Sigma}_k|) + C_n k$$

where n is the sample size, k is the order of the fitted model, $\hat{\Sigma}_k^2$ is an estimate of the white noise covariance matrix, and C_n is a sequence of specified constants (for AIC, $C_n = 2m^2/n$, for Hannan and Quinn's modification of BIC, $C_n = 2m^2(\ln \ln n)/n$, where m is the dimension of the data vector). A resampling scheme is proposed to estimate an improved penalty factor C_n. Conditional on the data, this procedure produces a consistent estimate of p. Simulation results support the effectiveness of this procedure when compared with some of the traditional order selection criteria. Comments are also made on the use of Yule–Walker as opposed to conditional least squares estimations for order selection.

1. INTRODUCTION

In this paper, we extend the results of Chen *et al.* (1993) for AR model selection using resampling methods to the multivariate problem. Let

Received June 23, 1993; revised July 1995.

AMS 1980 subject classification: 62M10.

Key words and phrases: multivariate autoregressive processes, order determination, AIC, Yule–Walker estimation, resampling.

* Research supported by NSF Grants DMS 9100392 and 9105745.

0047-259X/96 $18.00

$\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ be n consecutive observations from the m-variate AR(p) model,

$$\mathbf{X}_t - \Phi_1 \mathbf{X}_{t-1} - \cdots - \Phi_p \mathbf{X}_{t-p} = \mathbf{Z}_t \tag{1.1}$$

where $\Phi_1, ..., \Phi_p$ are real $m \times m$ matrices, $\det(I_m - \Phi_1 z - \cdots - \Phi_p z^p) \neq 0$ for $|z| \leqslant 1$, and $\{\mathbf{Z}_t\} \sim \text{IID}(\mathbf{0}, \Sigma)$ (i.e. $\{\mathbf{Z}_t\}$ is an independent and identically distributed sequence of random m-vectors with mean vector $\mathbf{0}$ and covariance matrix Σ) with Σ nonsingular. Most order selection criteria estimate the order of the model by minimizing an objective function of the form:

$$\delta(k) = \ln(|\hat{\Sigma}_k|) + C_n k, \tag{1.2}$$

where C_n is a sequence of prespecified constants and $\hat{\Sigma}_k$ is an estimate of Σ based on fitting an m-dimensional AR(k) model to the data ($|A|$ denotes the determinant of the matrix A). The estimated order, $\hat{p}$, is then defined as the minimizer of $\delta(k)$ over some suitable range of values. Typical values for the penalty factor are:

$$C(n, k) = \frac{2km^2}{n}, \qquad \text{AIC (Akaike (1973))}$$

$$C(n, k) = \frac{2(km^2 + m(m+1)/2)}{n - (km + m + 1)}, \qquad \text{AIC}_C \text{ (Hurvich and Tsai (1989), (1991))}$$

$$C(n, k) = \frac{2c(\log\log n)\, km^2}{n}, \qquad \text{(Hannan and Quinn (1979)).}$$

As in the univariate case, Hannan and Quinn's penalty factor produces a strongly consistent estimate of the true order p (see Hannan and Deistler, 1988).

If the true order of the model were known to be p, then we show (see Proposition 2.2) that there is a range of values of C_n for which $\delta(k)$ has its minimum at $k = p$. We call such penalty factors correct. In this paper we use a resampling scheme to approximate the range of correct penalty factors. The appeal of this method is that it produces a range of penalty factors which are correct for the simulated data. These test realizations are generated from candidate AR models–models which have been fitted to the original data. Since these penalty functions are correct for the simulated data from the candidate models and since the original model is believed to belong to the class of candidate models, it is reasonable to suppose the penalty factors will be nearly correct for the original data. This is borne out in Section 2 where we show that this procedure produces a weakly consistent estimate of the true order of the model and in Section 3 where we

show that perform extremely well in a simulation study. In Section 4, we discuss the effect of using Yule–Walker as opposed to conditional least squares estimates for order selection.

2. The Use of Resampling to Choose a Penalty Factor

In this section, we develop the necessary theory to establish weak consistency of a resampling procedure for model selection. Many of the results in this section parallel those for the univariate case given in Section 2 of Chen *et al.* (1993), however, the arguments used to derive them are substantially more difficult in two important respects. The first is the need for the strict monotonicity property of the determinant of the prediction error covariance matrices established in Proposition 2.1. The second is the delicate proof of the convergence of the bootstrapped multivariate Yule-Walker estimate conditional on the data.

Let $\{\mathbf{X}_t, t = 0, \pm 1, \ldots\}$ be the m-variate AR(p) process defined by (1.1). The estimated order, $\hat{p}$, based on observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$, is the minimizer of $\delta(k)$ where

$$\delta(k) = \ln(|\hat{\Sigma}_k|) + C_n k, \tag{2.1}$$

C_n is a sequence of as yet unspecified constants and $\hat{\Sigma}_k$ is the Yule–Walker estimate of Σ based on fitting an m-dimensional AR(k) model to the data.

For $k = 1, 2, \ldots$, let $\hat{\mathbf{X}}_{k+1} = \Phi_{k1}\mathbf{X}_k + \cdots + \Phi_{kk}\mathbf{X}_1$ denote the best linear predictor of $\mathbf{X}_{k+1}$ in terms of $\mathbf{X}_k, \ldots, \mathbf{X}_1$. The coefficient matrices $\Phi_{k1}, \ldots, \Phi_{kk}$ and the prediction error covariance matrix $\Sigma_k = E(\mathbf{X}_{k+1} - \hat{\mathbf{X}}_{k+1})(\mathbf{X}_{k+1} - \hat{\mathbf{X}}_{k+1})'$ are given by the Yule–Walker equations,

$$\begin{aligned} \Gamma(j) &= \Phi_{k1}\Gamma(j-1) + \cdots + \Phi_{kk}\Gamma(j-k), \qquad j = 1, \ldots, k, \\ \Sigma_k &= \Gamma(0) - \sum_{i=1}^{k} \Phi_{ki}\Gamma(k-i), \end{aligned} \tag{2.2}$$

where $\Gamma(h) := E[\mathbf{X}_h \mathbf{X}_0']$ is the matrix covariance function. Since $\Phi_p \neq 0_{m\times m}$ in (1.1) (the true order is p), we have for $k \geqslant p$

$$\mathbf{\Phi}_{ki} = \begin{cases} \Phi_i, & i \leqslant p, \\ 0_{m\times m}, & i > p, \end{cases} \tag{2.3}$$

and

$$\Sigma_k = \Sigma, \tag{2.4}$$

and for $k < p$

$$\Sigma_k \geqslant \Sigma, \tag{2.5}$$

where $A \geqslant B$ means the matrix $A - B$ is nonnegative definite and $0_{m \times m}$ is the $m \times m$ zero matrix. The Yule–Walker estimates $\hat{\Phi}_{k1}, ..., \hat{\Phi}_{kk}$ and $\hat{\Sigma}_k$ are obtained by replacing $\Gamma(h)$ by the sample matrix covariance function

$$\hat{\Gamma}(h) = \begin{cases} \frac{1}{n} \sum_{j=1}^{n-h} \mathbf{X}_{j+h} \mathbf{X}_j', & \text{if} \quad h \geqslant 0, \\ \hat{\Gamma}'(-h), & \text{if} \quad h < 0, \end{cases}$$

in equations (2.2).

PROPOSITION 2.1. *For an m-variate AR(p) process we have*

$$|\Sigma| = |\Sigma_p| < |\Sigma_j|$$

for $j = 0, ..., p-1$.

Proof. It follows from the definition of $\hat{\mathbf{X}}_{k+1}$ and Σ_k that

$$\Sigma_k \geqslant \Sigma_{k+1}$$

and hence

$$|\Sigma_k| \geqslant |\Sigma_{k+1}|.$$

So it suffices to show $|\Sigma_p| < |\Sigma_{p-1}|$. From the calculations on p. 422 of Brockwell and Davis (1991), it follows easily that

$$\Sigma_p = \Sigma_{p-1} - \Phi_{pp} \tilde{\Sigma}_{p-1} \Phi_{pp}'$$

where $\tilde{\Sigma}_{p-1}$ is the one-step error-covariance matrix of the best linear predictor of $\mathbf{X}_0$ in terms of $\mathbf{X}_1, ..., \mathbf{X}_{p-1}$. Writing $\Sigma_{p-1} = AA'$, where A is nonsingular, we have from the above relation that

$$\begin{aligned} 0 < |\Sigma_p \Sigma_{p-1}^{-1}| &= |I - A'^{-1} \Phi_{pp} \tilde{\Sigma}_{p-1} \Phi_{pp}' A^{-1}| \\ &= \prod_{i=1}^{m} (1 - \lambda_i) \end{aligned} \tag{2.6}$$

where $\lambda_1, ..., \lambda_m$ are the eigenvalues of $A'^{-1} \Phi_{pp} \tilde{\Sigma}_{p-1} \Phi_{pp}' A^{-1}$. Since $\Phi_{pp} = \Phi_p \neq 0$ and $\tilde{\Sigma}_{p-1}$ is positive definite, this matrix must have at least one nonzero eigenvalue. Consequently, the bound in (2.6) is strictly less than one. This completes the proof of the proposition. ∎

The following proposition determines the range of correct penalty factors (i.e., penalty factors such that the minimizer of $\delta(j, C_n)$ is p) when the true order of the model is known.

PROPOSITION 2.2. *Let* $\mathbf{X}_1, ..., \mathbf{X}_n$ *be observations from the* $AR(p)$ *model* (1.1) *and let* $K \geqslant p$ *be a fixed integer. Define,*

$$\alpha_n = \begin{cases} 0, & \text{for} \quad p = K, \\ \max\limits_{p+1 \leqslant l \leqslant K} \left\{ \dfrac{\ln(|\hat{\Sigma}_p|) - \ln(|\hat{\Sigma}_l|)}{l-p} \right\}, & \text{for} \quad p < K, \end{cases} \tag{2.7}$$

$$\beta_n = \begin{cases} +\infty, & \text{for} \quad p = 0, \\ \min\limits_{0 \leqslant l \leqslant p-1} \left\{ \dfrac{\ln(|\hat{\Sigma}_l|) - \ln(|\hat{\Sigma}_p|)}{p-l} \right\}, & \text{for} \quad p > 0, \end{cases} \tag{2.8}$$

where $\hat{\Sigma}_0 = \hat{\Gamma}(0)$. *Then as* $n \to \infty$,

(a) $\alpha_n \to 0$ *and* $\beta_n \to b > 0$ *a.s.*

(b) *Put*

$$\delta(k, C) = \ln(|\hat{\mathbf{\Sigma}}(k)|) + kC.$$

If $\alpha_n \leqslant \beta_n$, *then for any* $C_n \in (\alpha_n, \beta_n)$,

$$\delta(p, C_n) = \min_{0 \leqslant l \leqslant K} \{\delta(l, C_n)\}. \tag{2.9}$$

In other words, with this choice of penalty factor, the order of model (1.1) *will be correctly estimated by the minimizer of* (2.1).

Proof. (a) From Theorem 7.4.5 of Hannan and Deistler (1988), the Yule–Walker estimates are strongly consistent. Thus by (2.3)–(2.6), we have, for $p < K$

$$\alpha_n \xrightarrow{\text{a.s.}} \max_{p < j \leqslant K} \frac{\ln(|\Sigma_p|) - \ln(|\Sigma_j|)}{j-p} = 0$$

and from Proposition 2.1,

$$\beta_n \xrightarrow{\text{a.s.}} \min_{0 \leqslant j < p} \frac{\ln(|\Sigma_j|) - \ln(|\Sigma_p|)}{p-j} = b > 0, \quad p > 0.$$

The proof for the cases $p = K$ and $p = 0$ are immediate.

(b) Since $C_n \geqslant \alpha_n$, we have for $j > p$,

$$\begin{aligned}\ln(|\hat{\Sigma}_p|) - \ln(|\hat{\Sigma}_j|) + C_n(p-j) &< \ln(|\hat{\Sigma}_p|) - \ln(|\hat{\Sigma}_j|) + \frac{\ln(|\hat{\Sigma}_p|) - \ln(|\hat{\Sigma}_j|)}{j-p}(p-j) \\ &= 0,\end{aligned}$$

whence

$$\begin{aligned}\delta(j, C_n) &= \ln(|\hat{\Sigma}(j)|) + jC_n \\ &> \ln(|\hat{\Sigma}(p)|) + pC_n \\ &= \delta(p, C_n). \end{aligned} \tag{2.10}$$

On the other hand, since $C_n \leqslant \beta_n$, we have for $j < p$

$$\delta(p, C_n) - \delta(j, C_n) = \ln(|\hat{\Sigma}_p|) - \ln(|\hat{\Sigma}_j|) + C_n(p-j) < 0. \tag{2.11}$$

Combining (2.10) ant (2.11), we get (2.9). ∎

When the true order p of the model is known, the range of penalty factors C_n which leads to the correct model identification using (2.1) is nonempty, at least for large n. When p is unknown, we use a resampling procedure to generate test sequences from an AR(k) model for $k = 0, ..., K_1$, from which the above proposition may be applied to compute an interval of suitable values for C_n for each k. The intersection of these K_1 intervals then gives us a range from which to choose a penalty factor to be used in (2.1) applied to the original data. The fact that the intersection of these K_1 intervals is asymptotically nonempty is the content of the following proposition.

PROPOSITION 2.3. *In addition to the assumptions in Proposition* 2.2, *assume the noise vector has finite fourth moments. Define the residual sequence*

$$\hat{\mathbf{Z}}_t = \mathbf{X}_t - \hat{\boldsymbol{\Phi}}_{K1}\mathbf{X}_{t-1} - \cdots - \hat{\boldsymbol{\Phi}}_{KK}\mathbf{X}_{t-K}$$

for $t = K+1, ..., n$. *For any fixed integer* K_1 $(0 \leqslant K_1 \leqslant p)$, *let* $\{\mathbf{Y}_1^{(k)}, ..., \mathbf{Y}_n^{(k)}\}$ *be observations from the AR*(k) *model*

$$\mathbf{Y}_t^{(k)} - \hat{\boldsymbol{\Phi}}_{k1}\mathbf{Y}_{t-1}^{(k)} - \cdots - \hat{\boldsymbol{\Phi}}_{kk}\mathbf{Y}_{t-k}^{(k)} = \mathbf{Z}_t^*, \qquad t = k+1, ..., n; \quad k = 0, 1, ..., K_1,$$

where $\{\mathbf{Z}_t^*\}$ *is an iid sequence generated from the empirical distribution function (corrected to have mean zero) based on* $\{\hat{\mathbf{Z}}_t\}$. *(For* $k = 0$, $\mathbf{Y}_t^{(0)} = \mathbf{Z}_t^*$*). For each* $k = 0, 1, ..., K_1$, *let* $I_n^{(k)} = [\alpha_n^{(k)}, \beta_n^{(k)}]$ *denote the interval*

obtained when Proposition 2.2 *is applied to the series* $\{\mathbf{Y}_t^{(k)}\}$ *with* $p = k$. *Then for almost all sample sequences of* $\{\mathbf{X}_t\}$, *we have*

(a) $\alpha_n := \max_{0 \leqslant k \leqslant K_1} \{\alpha_n^{(k)}\} \xrightarrow{P_n} 0,$

(b) $\beta_n := \min_{0 \leqslant k \leqslant K_1} \{\beta_n^{(k)}\} \xrightarrow{P_n} b \geqslant 0,$

where $\xrightarrow{P_n}$ *denotes convergence in probability conditional on* $\mathbf{X}_1, ..., \mathbf{X}_n$. *In particular,* $I_n = \bigcap_{k=0}^{K_1} I_n^{(k)}$ $(= [\alpha_n, \beta_n]$ *if* $\alpha_n \leqslant \beta_n)$ *converges in conditional probability to a nonempty set.*

Proof. Since the argument is nearly identical to the proof given for Proposition 2.2 in Chen, et al. (1993), we only give a sketch of the argument. It suffices to show that

$$\alpha_n^{(k)} \xrightarrow{P_n} 0 \qquad \text{and} \qquad \beta_n^{(k)} \xrightarrow{P_n} b_k \geqslant 0. \tag{2.12}$$

Let $\hat{\Gamma}^*(h) := (1/n) \sum_{j=1}^{n} \mathbf{Y}_{t+h}^{(k)} \mathbf{Y}_t^{(k)'}$ denote the sample covariance matrix function of the $\mathbf{Y}_t^{(k)}$'s. We begin by showing that the conditional covariance matrix of $\Gamma^*(h)$, denoted by $\mathrm{Cov}_n(\hat{\Gamma}^*(h))$, convergences to 0, a.s. By Remark 2 on p. 424 of Brockwell and Davis (1991), the Yule–Walker estimates produce a causal model so that $\{\mathbf{Y}_t^{(k)}\}$ has the representation

$$\mathbf{Y}_t^{(k)} = \sum_{j=0}^{\infty} \hat{\Psi}_j \mathbf{Z}_{t-j}^*$$

where the $\hat{\Psi}_j$'s are the coefficient matrices in the power series expansion of the matrix function $(1 - \hat{\Phi}_{k1} z - \cdots - \hat{\Phi}_{kk} z^k)^{-1}$ on $|z| \leqslant 1$. The conditional mean of $\hat{\Gamma}^*(h)$ is given by

$$\Gamma^*(h) := E_n(\mathbf{Y}_{t+h}^{(k)} \mathbf{Y}_t^{(k)'}) = \sum_{j=0}^{\infty} \hat{\Psi}_{j+h} \Sigma^* \hat{\Psi}_j' \tag{2.13}$$

where $E_n(\cdot)$ denotes expectation relative to P_n and Σ^* is the sample covariance matrix of $\{\mathbf{Z}_t\}$. The strong consistency of the Yule–Walker estimates implies that for almost all sample paths, $\hat{\Phi}_k \to \Phi_k$ and $\hat{\Sigma}_k \to \Sigma_k$. It follows that $\hat{\Psi}_j \to \Psi_j$ as $n \to \infty$ (Ψ_j are the coefficient matrices in the expansion of the matrix function $(1 - \Phi_{k1} z - \cdots - \Phi_{kk} z^k)^{-1}$). Now for $l = 0, 1, ...,$ write

$$\hat{\Psi}_l = [\hat{\Psi}_l(i, j)]_{i,j=1}^{m}.$$

It is easy to see that there exist constants $C > 0$ and $\tau < 1$, depending on the sample path, such that

$$|\hat{\Psi}_l(i, j)| \leqslant C r^l \tag{2.14}$$

for $l = 0, 1, \ldots$ and n large. Similarly, one can show that $\Sigma^* \to \Sigma_k$ which combined with (2.13) and (2.14) yields

$$\Gamma^*(h) \to \sum_{j=0}^{\infty} \Psi_{j+h} \Sigma_k \Psi_j' \tag{2.15}$$

as $n \to \infty$. Using these results together with the relations established in the proof of Theorem 6, p. 210 of (Hannan (1970)), we have

$$\limsup_{n \to \infty} n \operatorname{Cov}_n(\hat{\Gamma}^*(h)) < \infty \qquad \text{a.s.}$$

This implies

$$\operatorname{Cov}_n(\hat{\Gamma}^*(h)) \to 0 \qquad \text{a.s.}$$

which together with (2.15) yields

$$\hat{\Gamma}^*(h) \xrightarrow{P_n} \sum_{j=0}^{\infty} \Psi_{j+h} \Sigma_k \Psi_j'.$$

The weak consistency of the sample act implies that the Yule–Walker estimates, in fitting an AR(j) model to the data $\{\mathbf{Y}_t^{(k)}\}$, are also weakly consistent relative to P_n. The limits in (2.12) now follow using the argument given for Proposition 2.1. ▌

THEOREM 2.4. *Under the assumptions of Proposition* 2.3, *let* $I_n = [\alpha_n, \beta_n]$ *and suppose* $\Phi_{jj} \neq \mathbf{0}$ *for* $j = 1, ..., K_1$. *Define*

$$C_n = \begin{cases} \alpha_n + \dfrac{c\beta_n \ln(n)}{n}, & \text{if} \quad \alpha_n < \beta_n, \\ \dfrac{c\beta_n \ln n}{n}, & \text{otherwise}, \end{cases}$$

where $c > 0$ *is any constant such that* $C_n \in I_n$. *If* $\hat{p}_n$ *is the minimizer of* $\delta(k, C_n)$ *for* $0 \leqslant k \leqslant K$, *then for almost all sample sequences of* $\{\mathbf{X}_t\}$,

$$\hat{p}_n \xrightarrow{P_n} p \tag{2.17}$$

as $n \to \infty$.

Proof. Since for almost all sample paths,

$$C_n \xrightarrow{P_n} 0 \qquad \text{and} \qquad \frac{nC_n}{\log \log n} \xrightarrow{P_n} \infty,$$

the theorem follows at once from Theorem 5.5.1 of Hannan and Deistler (1988). ▮

Remark 1. In order to implement our resampling procedure for determining C_n using Theorem 2.4, we must first specify $K_1 \leqslant p$ such that $\Phi_{jj} \neq \mathbf{0}$ for $j = 1, ..., K_1$. Ideally, we would like to optimize over as large a class of test models as possible in order to ensure that a model close to the true model is included in our test set. This requires a large K_1. If our initial choice for K_1 happens to be bigger than p, then it is likely that the set I_n will be empty. In this case, the value of K_1 is reduced in steps of size 1 until a nonempty I_n is achieved.

Remark 2. While we have assumed in Proposition 2.3 that the test series $\{\mathbf{Y}_t^{(k)}\}$ has been generated from the candidate AR(k) model, this is not necessary. If instead, $\{\mathbf{Y}_1^{(k)}, ..., \mathbf{Y}_n^{(k)}\}$ were generated from the model

$$\mathbf{Y}_t^{(k)} - A_{k1}\mathbf{Y}_{t-1}^{(k)} - \cdots - A_{kk}\mathbf{Y}_{t-k}^{(k)} = \mathbf{Z}_t^*$$

where $A_{k1}, ..., A_{kk}$ are prespecified coefficient matrices, then the conclusions of Proposition 2.3 and Theorem 2.4 would remain unchanged. Two potential advantages of generating the test sequences in this way are that the condition $K_1 \leqslant p$ is no longer required and that a coefficient matrix A_{kk} very different from the zero matrix increases the likelihood that the set $I_n = [\alpha_n, \beta_n]$ is nonempty.

3. Implementation and Simulation

The resampling procedure for order selection of multivariate AR models is basically the same as for the univariate case. Assume that $\mathbf{X}_1, ..., \mathbf{X}_n$ are observations from a multivariate AR(p) process defined as in (1.1). The order selection procedure is implemented as follows:

Step 1. Choose a fixed integer K which is believed to be greater than the true order p and compute the Yule–Walker estimates $\hat{\Phi}_{K1}, ..., \hat{\Phi}_{KK}, \hat{\Sigma}_K$ from the observed data, $\{\mathbf{X}_t\}_{t=1}^n$. The residual sequence is given by

$$\hat{\mathbf{Z}}_t = \mathbf{X}_t - \hat{\Phi}_{K1}\mathbf{X}_{t-1} - \cdots - \hat{\Phi}_{KK}\mathbf{X}_{t-K}$$

for $t = K+1, ..., n$. Center the residuals by subtracting off the sample mean $(1/n-K)\sum_{t=K+1}^n \hat{\mathbf{Z}}_t$. For simplicity, we use the same notation $\{\hat{\mathbf{Z}}_t\}_{K+1}^n$ for the centered residuals.

Step 2. Compute the Yule–Walker estimates $\hat{\Phi}_k, \hat{\Sigma}_k, k = 0, 1, ..., K$ from the observed data, $\{\mathbf{X}_t\}_{t=1}^n$.

Step 3. Choose a positive integer $K_1 \leqslant K$. For $k = 0, ..., K_1$ generate observations $\mathbf{Y}_1^{(k)}, ..., \mathbf{Y}_n^{(k)}$ from the model

$$\mathbf{Y}_t^{(k)} - \hat{\boldsymbol{\Phi}}_{k1}\mathbf{Y}_{t-1}^{(k)} - \cdots - \hat{\boldsymbol{\Phi}}_{kk}\mathbf{Y}_{t-k}^{(k)} = \mathbf{Z}_t^*$$

where $\{\mathbf{Z}_t^*\}$ is an IID sequence sampled from the centered residuals, $\{\hat{\mathbf{Z}}_t\}_{K+1}^n$. The case $k = 0$ corresponds to $\mathbf{Y}_t^{(0)} = \mathbf{Z}_t^*$.

Step 4. For $k = 0, ..., K_1$, compute the Yule–Walker estimate of the innovation covariance matrix in fitting an AR(j) model to $\{\mathbf{Y}_t^{(k)}\}_{t=1}^n$ for $j = 0, ..., K$. Denote this estimate by $\hat{\Sigma}_j(k)$.

Step 5. For $k = 0, ..., K_1$ compute

$$\alpha_n^{(k)} = \begin{cases} 0, & \text{if } k = K, \\ \max\limits_{k < j \leqslant K} \dfrac{\log|\hat{\Sigma}_k(k)| - \log|\hat{\Sigma}_j(k)|}{j-k}, & \text{if } k < K, \end{cases}$$

and

$$\beta_n^{(k)} = \begin{cases} \infty, & \text{if } k = 0, \\ \min\limits_{0 \leqslant j < k} \dfrac{\log|\hat{\Sigma}_j(k)| - \log|\hat{\Sigma}_k(k)|}{k-j}, & \text{if } k > 0. \end{cases}$$

Step 6. Compute

$$\alpha_n = \max_{0 \leqslant l \leqslant K_1} \{\alpha_n^{(k)}\}$$

and

$$\beta_n = \min_{0 \leqslant k \leqslant K_1} \{\beta_n^{(k)}\}.$$

Step 7. If $\alpha_n < \beta_n$ set

$$C_n = \alpha_n + \frac{c(\log n)\,\beta_n}{n}$$

where $c > 0$ is such that $C_n \leqslant \beta_n$. If $\alpha_n \geqslant \beta_n$, then reduce the value of K_1 by 1 and return to Step 6.

Step 8. The estimated order $\hat{p}$ is defined to be the minimizer of

$$\delta(k, C_n) = \log|\hat{\Sigma}_k| + kC_n$$

for $0 \leqslant k \leqslant K$.

To reduce sampling variability, it is often beneficial to generate many replicates of the test series $\{\mathbf{Y}_t^{(k)}\}_{t=1}^{n}$ in Step 4. The computed values of $\alpha_n^{(k)}$ and $\beta_n^{(k)}$ in Step 5 are then replaced by their respective averages over the replications.

We also considered a modification of this procedure mentioned in Remark 2 of the preceding section. The modification occurs in Step 4 where the test series $\{\mathbf{Y}_t^{(k)}\}$ is generated from an AR(k) model with a prespecified sequence of parameter vectors $(A_{k1}, ..., A_{kk})'$.

In our simulation study, we compared our proposed procedure and its modification with the following four well known order selection criteria:

$$\text{AIC} \qquad \log|\hat{\Sigma}_k| + \frac{2km^2}{n}$$

$$\text{AICC} \qquad \log|\hat{\Sigma}_k| + \frac{2(km^2+1)}{n-(km^2+2)}$$

$$\text{H \& Q} \qquad \log|\hat{\Sigma}_k| + \frac{(2\log\log n)\, km^2}{n}$$

$$\text{BIC} \qquad n\log[|\hat{\Sigma}_k| + m] + (m^2k + m(m+1)/2)\log n$$

where $\hat{\mathbf{\Phi}}$ and $\hat{\Sigma}_k$ are the Yule–Walker estimates of the coefficients and innovation covariance matrix in fitting an m-variate AR(k) model to the data. Note that the definition of AICC here (see Brockwell and Davis (1991), p. 432) differs from the AIC_C defined in Section 1. However, in the simulation results of Hurvich and Tsai (1991), AICC outperformed AIC_C in terms of the frequency of correct identifications of the true order of the model, and hence we have used the former in our study.

We generated 100 sample paths of various lengths from each of the following AR models:

$$\mathbf{X}_t = \begin{pmatrix} -1.0 & 0.96 \\ -1.5 & 1.4 \end{pmatrix} \mathbf{X}_{t-1} + \mathbf{Z}_t \tag{3.1}$$

$$\mathbf{X}_t = \begin{pmatrix} 0.5 & -0.3 \\ 0.2 & 0.65 \end{pmatrix} \mathbf{X}_{t-1} + \begin{pmatrix} -0.5 & 0.3 \\ 0.0 & -0.4 \end{pmatrix} \mathbf{X}_{t-2} + \mathbf{Z}_t \tag{3.2}$$

$$\mathbf{X}_t = \begin{pmatrix} -0.17 & 0.14 \\ -0.19 & -0.1 \end{pmatrix} \mathbf{X}_{t-1} + \begin{pmatrix} -0.2 & 0.12 \\ 0.22 & -0.25 \end{pmatrix} \mathbf{X}_{t-2} + \begin{pmatrix} 0.0 & 0.4 \\ 0.15 & -0.37 \end{pmatrix} \mathbf{X}_{t-3} + \mathbf{Z}_t \tag{3.3}$$

where $\{\mathbf{Z}_t\}$ is an IID sequence of $N(\mathbf{0}, \Sigma)$ random variables with

$$\Sigma = \begin{pmatrix} 1.0 & -0.08 \\ -0.08 & 1.0 \end{pmatrix}.$$

The AR(1) and AR(2) models were used by Hurvich and Tsai (1991) in their simulation study. In all of our simulations we took $K = 10$, $K_1 = 2$, and $c = 5.0$. The bounds $\alpha_n^{(k)}$ and $\beta_n^{(k)}$ were computed as an average based on 50 replicates of the test series. For the modified procedure (MDC), the parameter vectors were

$$A_{11} = \begin{pmatrix} -0.6 & 0.0 \\ 0.0 & 0.5 \end{pmatrix},$$

and

$$A_{21} = \begin{pmatrix} 1.4 & 0.0 \\ 0.0 & 0.4 \end{pmatrix}, \qquad A_{22} = \begin{pmatrix} 0.49 & 0.0 \\ 0.0 & 0.3 \end{pmatrix}.$$

The frequencies of the estimated orders for each of the 6 criteria are summarized in Tables I–V (the method described above and its modification are listed as DC and MDC, respectively).

As in the univariate case, the modified procedure (MDC) generally outperformed the other procedures, although the margin was not as great as in the univariate case. In a sense, order selection for multivariate AR's is easier than in the univariate case since an increase in the order of the model from k to $k+1$ adds m^2 more parameters to the model. A more delicate model identification problem is to identify both the order of the model and possible constraints on the parameters.

TABLE I

Frequencies of Estimated Order in 100 Replications from the AR(1) Model Given by (3.1) with Sample Sizes 50 and 100 (in Parentheses)

	Estimated order					
Criterion	0	1	2	3	4	5–10
AIC	0 (0)	96 (88)	0 (8)	4 (0)	0 (2)	0 (0)
AICC	0 (0)	98 (96)	0 (4)	2 (0)	0 (0)	0 (0)
H & Q	0 (0)	98 (98)	0 (2)	2 (0)	0 (0)	0 (0)
BIC	0 (0)	99 (100)	1 (0)	0 (0)	0 (0)	0 (0)
DC	0 (0)	100 (97)	0 (1)	0 (1)	0 (1)	0 (0)
MDC	0 (0)	100 (100)	0 (0)	0 (0)	0 (0)	0 (0)

TABLE II

Frequencies of Estimated Order in 100 Replications from the AR(2) Model Given by (3.2) with Sample Sizes 50 and 100 (in Parentheses)

	Estimated order					
Criterion	0	1	2	3	4	5–10
AIC	0 (0)	0 (0)	89 (88)	7 (11)	2 (0)	2 (1)
AICC	0 (0)	2 (0)	98 (95)	0 (5)	0 (0)	0 (0)
H & Q	0 (0)	2 (0)	96 (99)	2 (1)	0 (0)	0 (0)
BIC	0 (0)	4 (0)	96 (100)	0 (0)	0 (0)	0 (0)
DC	0 (0)	1 (0)	97 (99)	2 (1)	0 (0)	0 (0)
MDC	0 (0)	6 (0)	94 (100)	0 (0)	0 (0)	0 (0)

TABLE III

Frequencies of Estimated Order in 100 Replications from the AR(3) Model Given by (3.3) with Samples Sizes 50, 100 (in Parentheses), and 200 (in Square Brackets)

	Estimated order					
Criterion	0	1	2	3	4	5–10
AIC	2 (0) [0]	2 (0) [0]	6 (0) [0]	62 (88) [92]	16 (8 [8]	12 (4) [0]
AICC	16 (0) [0]	8 (2) [0]	14 (0) [0]	62 (96) [96]	0 (2) [4]	0 (0) [0]
H & Q	16 (0) [0]	4 (2) [0]	12 (0) [0]	62 (98) [99]	4 (0) [1]	2 (0) [0]
BIC	54 (8) [0]	6 (4) [0]	8 (4) [0]	32 (84) [100]	0 (0) [0]	0 (0) [0]
DC	12 (2) [0]	0 (2) [0]	10 (0) [0]	66 (88) [99]	8 (6) [1]	4 (2) [0]
MDC	12 (0) [0]	2 (2) [0]	12 (0) [0]	63 (92) [100]	7 (6) [0]	4 (0) [0]

TABLE IV

Frequencies of Correct Order Selection in 100 Replications from the Models (3.1)–(3.4) Using DC

	Sample size		
Model	50	100	200
AR(1)	100	97	94
AR(2)	97	99	100
AR(3)	68	88	99

TABLE V

Frequencies of Correct Order Selection in 100 Replications from the Models (3.1)–(3.4) Using MDC

	Sample size		
Model	50	100	200
AR(1)	100	100	100
AR(2)	94	100	100
AR(3)	63	92	100

4. Yule–Walker vs Conditional Least-Squares

In the simulation results of the preceding section, the parameters were all estimated using the Yule–Walker equations. We also compared these results with estimates based on conditional least squares (see Hurvich & Tsai (1991) and Lütkepohl (1991)). For small to moderate sample sizes, the order selection criteria performed consistently better with the Yule–Walker estimates. This difference is even more pronounced in the univariate case. (See Chen et al. (1993) where Yule–Walker estimation was compared with both Burg and maximum likelihood estimation.)

To demonstrate the difference in performance between the two estimates, we generated 500 time series of length 30 from the model (3.1),

$$\mathbf{X}_t = \begin{pmatrix} -1.0 & 0.96 \\ -1.5 & 1.4 \end{pmatrix} \mathbf{X}_{t-1} + \mathbf{Z}_t,$$

where $\{\mathbf{Z}_t\}$ is an IID sequence of $N(\mathbf{0}, \Sigma)$ random vectors. An AR(6) model was fitted to each series using both Yule–Walker and conditional least squares estimation. The 24×1 parameter vector for the model is given by

$$\mathbf{a}' = (-1.0, .96, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -1.5, 1.4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0).$$

Figures 1 and 2 show the boxplots of the Yule–Walker and least squares estimates respectively of the components of the vector $\mathbf{a}$. One can clearly see that the error bars of the Yule–Walker estimates are systematically shorter than those of the least squares estimates. The Yule–Walker estimators appear to be better behaved than conditional least squares estimators for fitting over-parameterized autoregressive models, giving generally smaller values for the coefficients at lags greater than the true order of the model. Accurate fitting of over-parameterized models is an important ingredient in model selection.

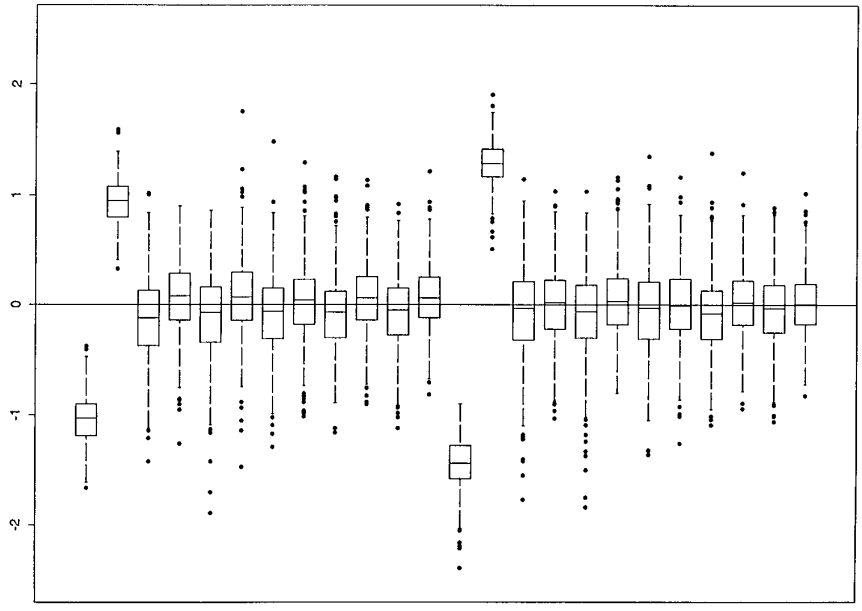

FIG. 1. AR(6) coefficients fitted by Yule–Walker. The true model is the AR(1) (3.1). Box plots are based on 500 replicates.

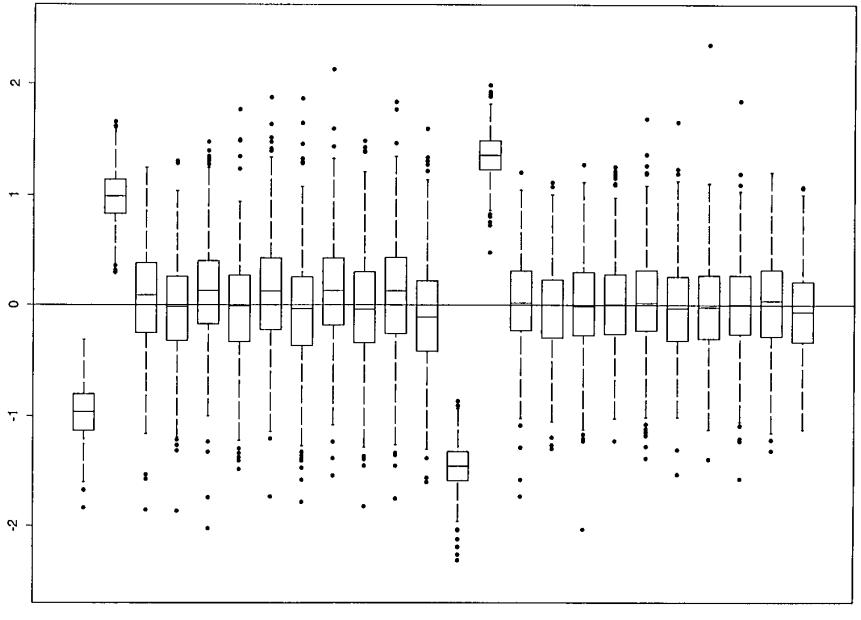

FIG. 2. AR(6) coefficients fitted by least squares. The true model is the AR(1) (3.1). Box plots are based on 500 replicates.

References

Akaike, H. (1969). Fitting autoregressive models for prediction. *Ann. Inst. Statist. Math.* **21** 243–247.

Brockwell, P. J., and Davis, R. A. (1991). *Time Series*: *Theory and Methods*. Springer-Verlag, New York.

Chen, C., Davis, R. A., Brockwell, P. J., and Bai, Z. D. (1993). Order determination for autoregressive processes using resampling methods. *Statistica Sinica* **3** 481–500.

Hannan, E. J. (1970). *Multiple Time Series*. J. Wiley, New York.

Hannan, E. J., and Deistler, M. (1988). *The Statistical Theory of Linear Systems*. Wiley, New York.

Hannan, E. J., and Quinn, B. G. (1979). The determination of the order of an autoregression. *J. R. Statist. Soc. B.* **41** (2) 190–195.

Hurvich, C. M., and Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika* **76** (2) 297–307.

Hurvich, C. M., and Tsai, C. L. (1991). A corrected Akaike information criterion for vector autoregresive model selection. *J. Time Series Anal.* **14** 271–280.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.